

Geuvadis

Allele-specific alternative splicing

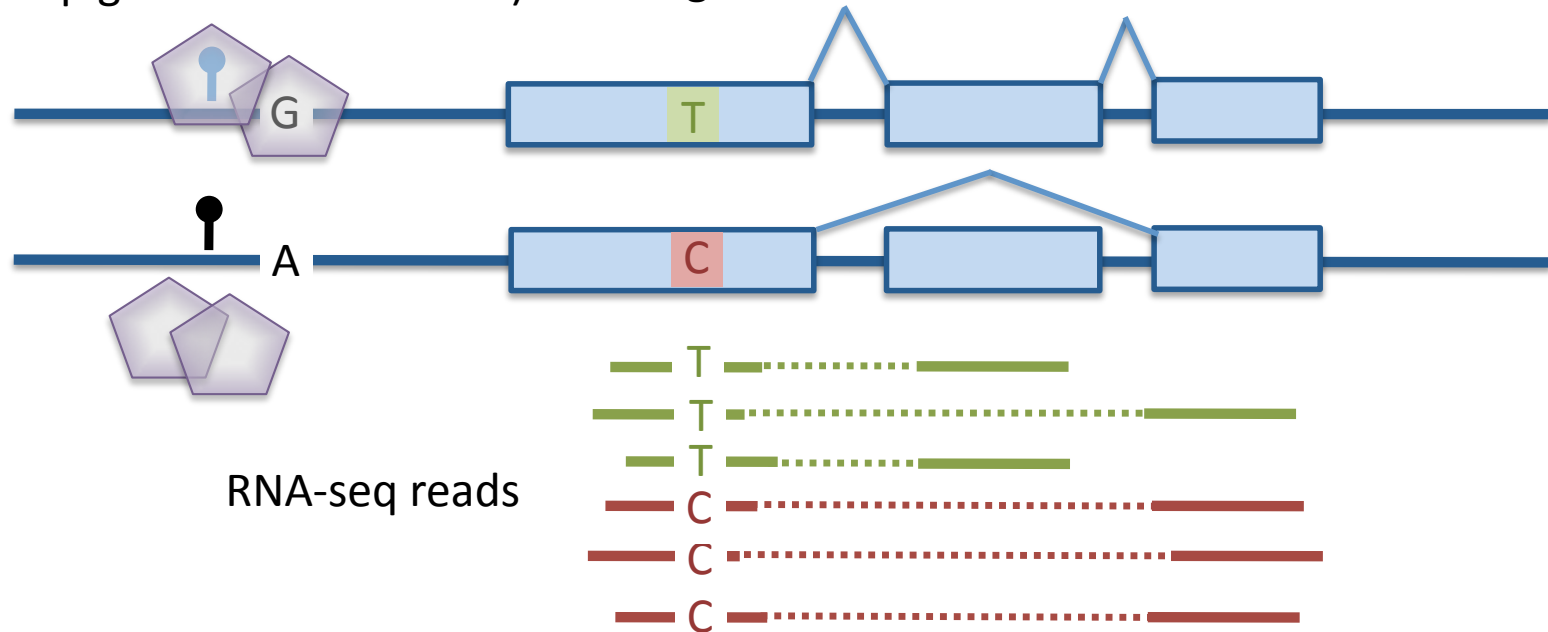
Tuuli Lappalainen

October 11, 2012

Allele-specific Alternative Splicing (ASAS)

coding or noncoding *cis* asQTL
(or epigenetic modification)

coding SNP



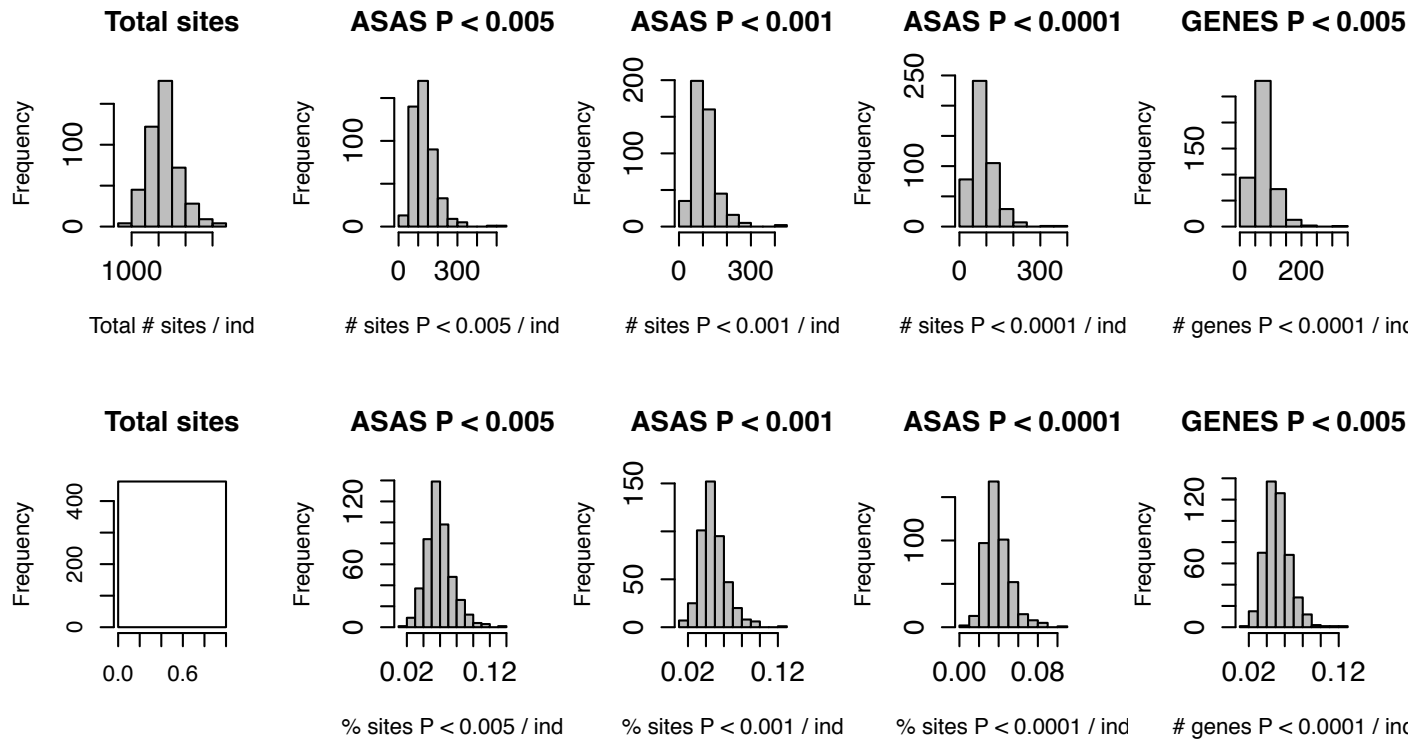
Exon counts

T-reads	3	2	1
C-reads	3	0	3

ASAS pipeline

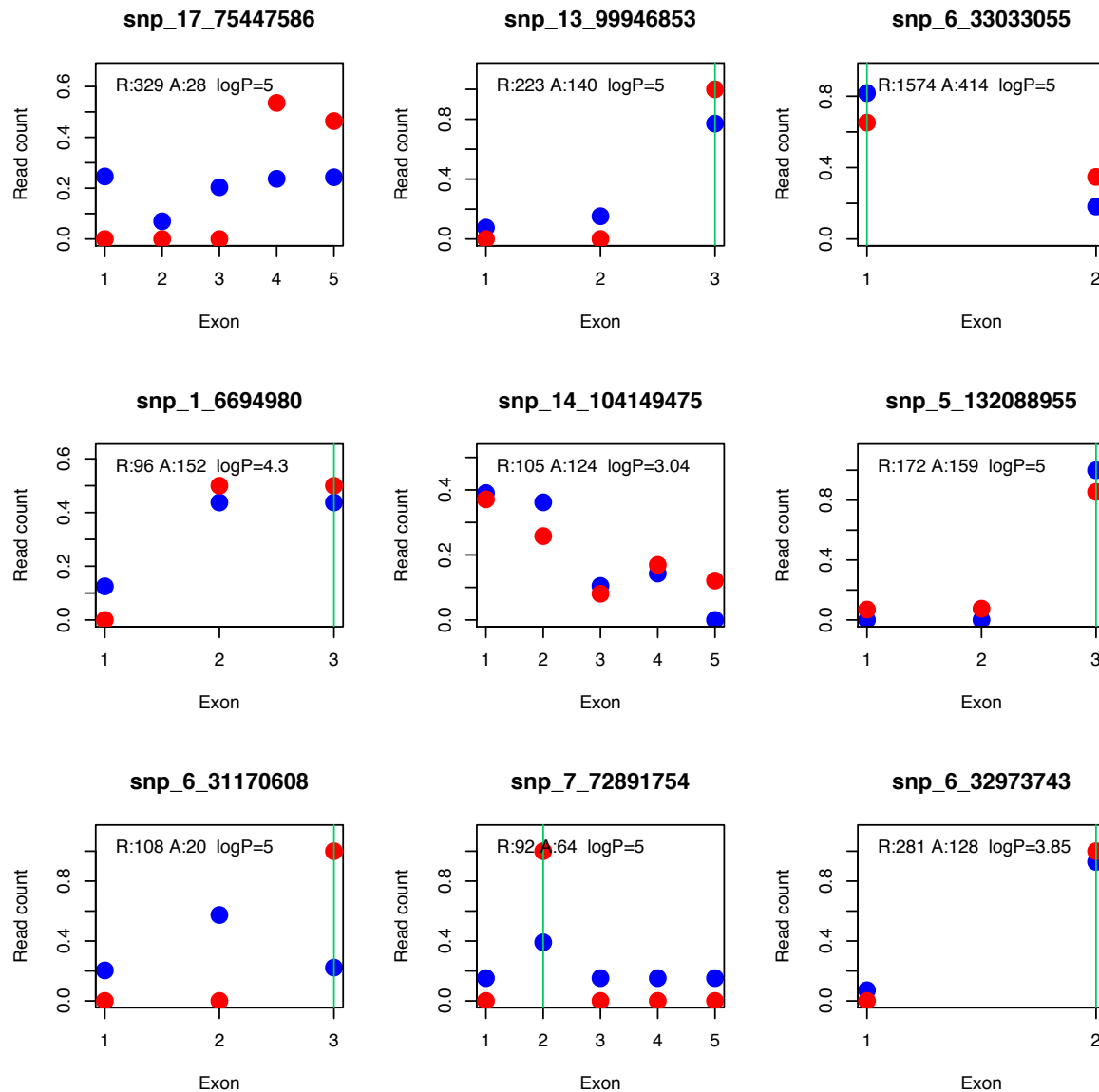
- For each individual, get all heterozygous sites that are covered by ≥ 20 RNAseq reads (or whatever limit, but you need more reads than in ASE)
 - use ASE pipeline results for this filtering
- Search the bam file for these sites using the pysam package
 - extract all the reads and their mates that overlap
 - separate them to reads with REF or ALT allele
 - print them out as pseudo-sam file
- For each site, calculate the number of REF and ALT read overlaps in exons
- Fisher test for exon counts
- Quantitative measure analogous to ASE allelic ratio – one number for each het site:
max of the ratios per exon, normalized by total allelic ratio :
$$\left[\text{abs} \left(0.5 - \frac{\text{REF}}{\text{REF} + \text{ALT}} \right) \right] / \left[\frac{\text{sum}(\text{REF})}{\text{sum}(\text{REF} + \text{ALT})} \right]$$
- Combine output
- The whole pipeline can be run in ~ 30 h for 667 Geuvadis samples

Depth of the data

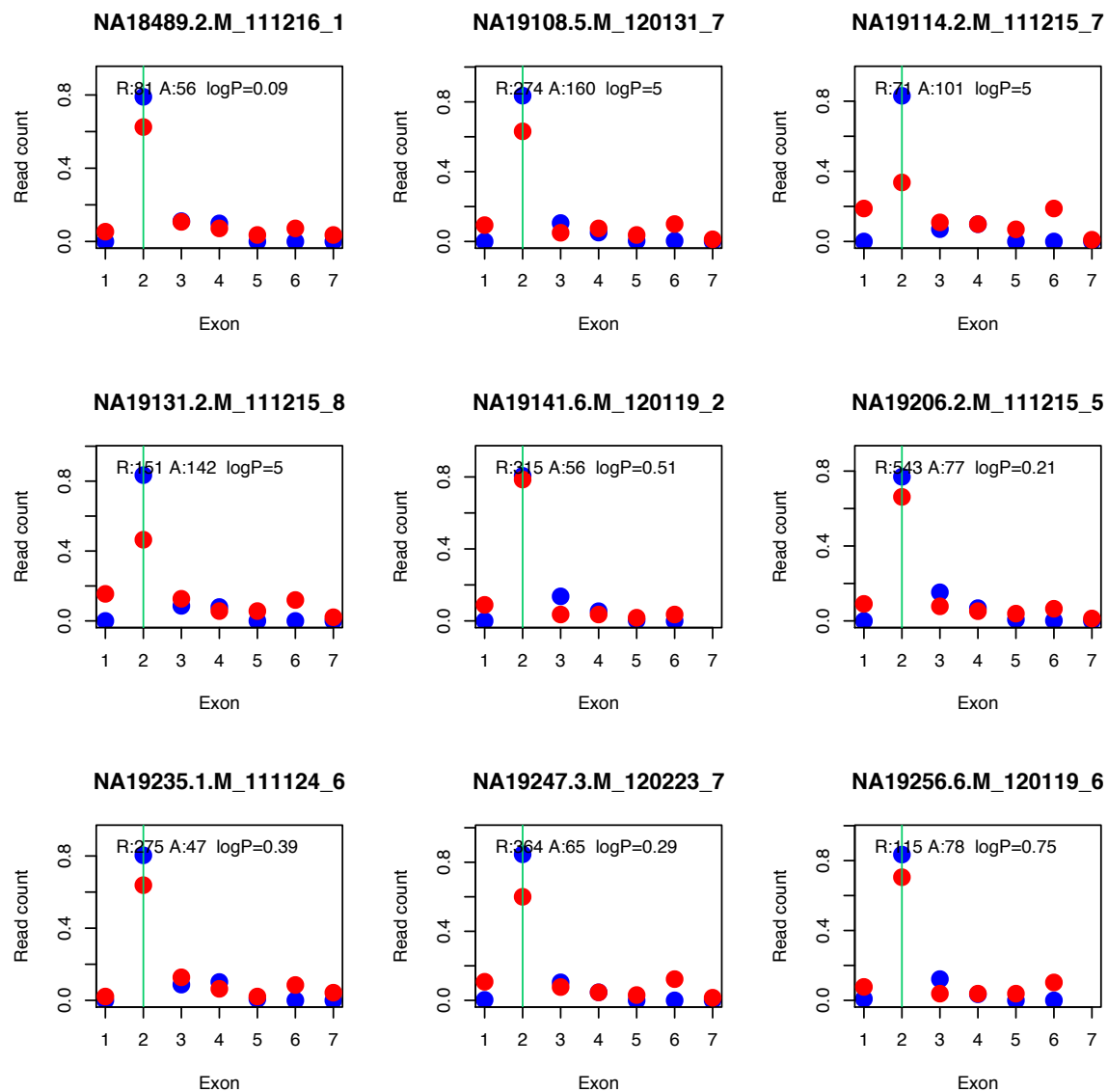


STATISTIC	MIN	MEDIAN	MAX
TOTAL SITES	740	2135	4443
SITES p<0.005	13	121	509
GENES	546	1415	2660
GENES w SITES p<0.005	13	74	326

Example: significant sites

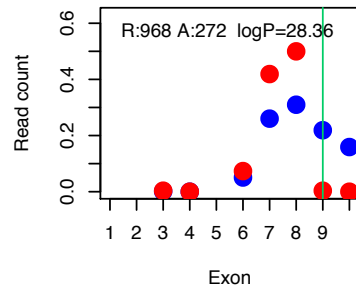


Example: one site, some with ASAS

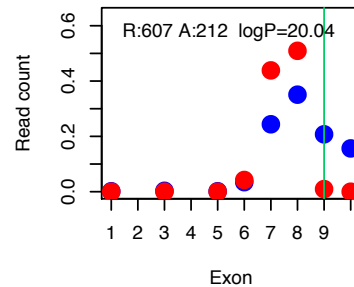


Example: splice site variant

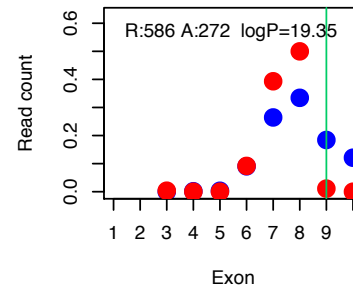
HG00110.2.M_120131_2



HG00111.2.M_111215_4

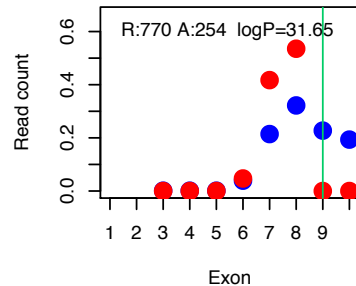


HG00114.1.M_120209_3

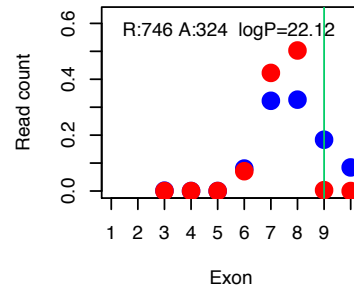


- More analysis on ASAS effects of annotated splice variants
- Absence of ASAS doesn't prove absence of splicing effect

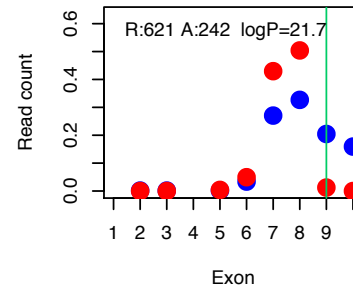
HG00116.2.M_120131_1



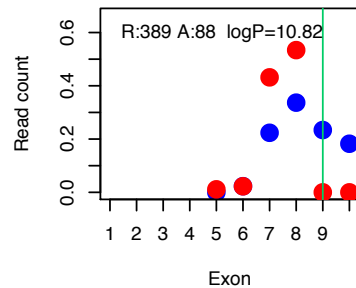
HG00117.1.M_111124_2



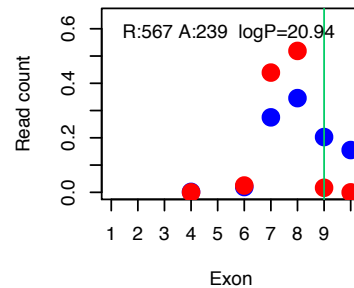
HG00117.1.M_120209_1



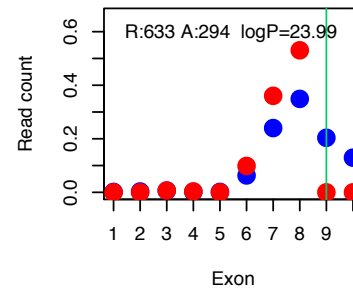
HG00117.3.M_120202_6



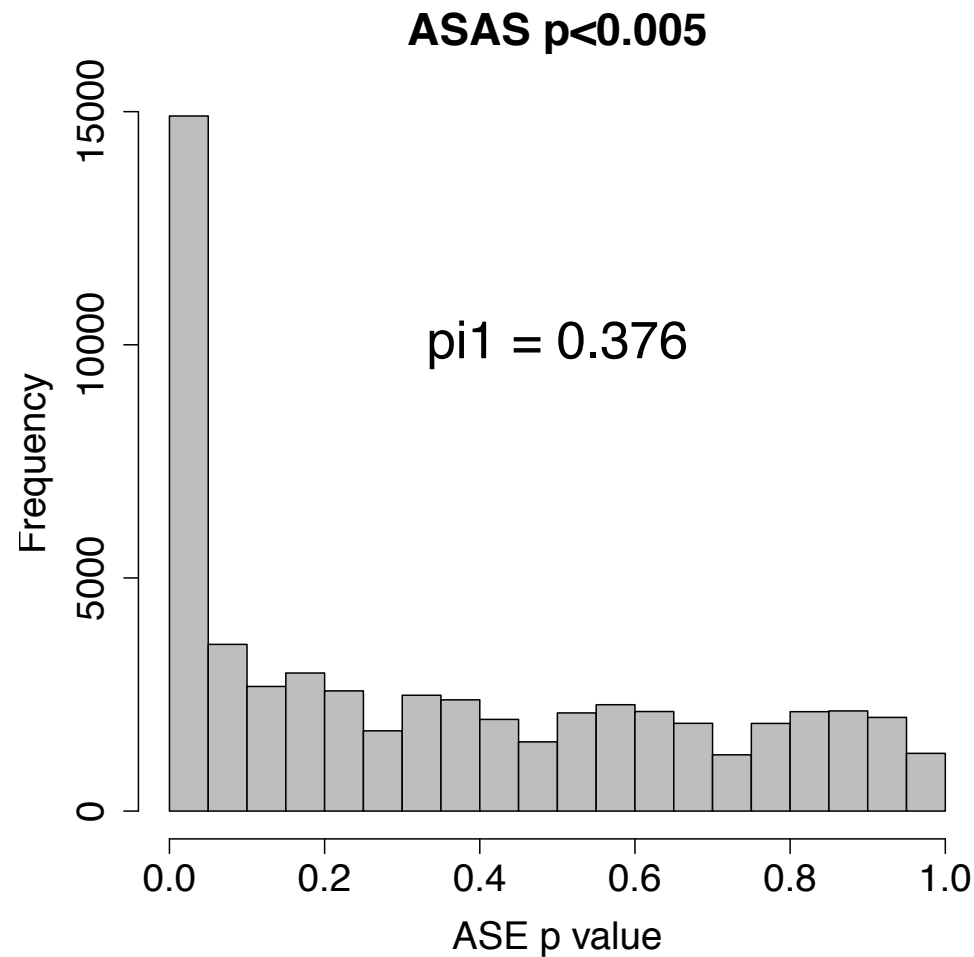
HG00117.7.M_120219_4



HG00123.4.M_120208_7



ASAS and ASE sharing



- Splicing changes can cause ASE in the exons that are affected
- ...or underlying genetic causes can be shared
- further analysis needed